
Sequence Listing was accepted.

If you need help call the Patent Electronic Business Center at (866) 217-9197 (toll free).

Reviewer: Saleem, Syed (ASRC)

Timestamp: [year=2010; month=7; day=28; hr=11; min=0; sec=30; ms=525;]

Validated By CRFValidator v 1.0.3

Application No: 10522341 Version No: 3.0

Input Set:

Output Set:

Started: 2010-07-22 15:40:42.953

Finished: 2010-07-22 15:40:50.330

Elapsed: 0 hr(s) 0 min(s) 7 sec(s) 377 ms

Total Warnings: 126

Total Errors: 0
No. of SeqIDs Defined: 179

Actual SeqID Count: 179

Error code	Error Description
W 213	Artificial or Unknown found in <213> in SEQ ID (3)
W 213	Artificial or Unknown found in <213> in SEQ ID (4)
W 402	Undefined organism found in <213> in SEQ ID (13)
W 402	Undefined organism found in <213> in SEQ ID (14)
W 402	Undefined organism found in <213> in SEQ ID (15)
W 402	Undefined organism found in <213> in SEQ ID (16)
W 213	Artificial or Unknown found in <213> in SEQ ID (49)
W 213	Artificial or Unknown found in <213> in SEQ ID (50)
W 213	Artificial or Unknown found in <213> in SEQ ID (51)
W 213	Artificial or Unknown found in <213> in SEQ ID (52)
W 213	Artificial or Unknown found in <213> in SEQ ID (53)
W 213	Artificial or Unknown found in <213> in SEQ ID (54)
W 213	Artificial or Unknown found in <213> in SEQ ID (55)
W 213	Artificial or Unknown found in <213> in SEQ ID (56)
W 213	Artificial or Unknown found in <213> in SEQ ID (57)
W 213	Artificial or Unknown found in <213> in SEQ ID (58)
W 213	Artificial or Unknown found in <213> in SEQ ID (69)
W 213	Artificial or Unknown found in <213> in SEQ ID (70)
W 213	Artificial or Unknown found in <213> in SEQ ID (71)
W 213	Artificial or Unknown found in <213> in SEQ ID (72)

Input Set:

Output Set:

Started: 2010-07-22 15:40:42.953 **Finished:** 2010-07-22 15:40:50.330

Elapsed: 0 hr(s) 0 min(s) 7 sec(s) 377 ms

Total Warnings: 126
Total Errors: 0

No. of SeqIDs Defined: 179

Actual SeqID Count: 179

Err	or code	Error Description
W	213	Artificial or Unknown found in <213> in SEQ ID (73)
W	213	Artificial or Unknown found in <213> in SEQ ID (74)
W	213	Artificial or Unknown found in <213> in SEQ ID (75)
W	213	Artificial or Unknown found in <213> in SEQ ID (76) This error has occured more than 20 times, will not be displayed

```
<211> 1284
<212> DNA
<213> Escherichia coli
<220>
<221> CDS
<222> (1)..(1281)
<223> coding for cytosine deaminase (codA)
<400> 1
gtg tcg aat aac gct tta caa aca att att aac gcc cgg tta cca ggc
                                                                   48
Val Ser Asn Asn Ala Leu Gln Thr Ile Ile Asn Ala Arg Leu Pro Gly
                                     10
qaa qaq qqq ctq tqq caq att cat ctq caq qac qqa aaa atc aqc qcc
Glu Glu Gly Leu Trp Gln Ile His Leu Gln Asp Gly Lys Ile Ser Ala
                                 25
att gat gcg caa tcc ggc gtg atg ccc ata act gaa aac agc ctg gat
                                                                   144
Ile Asp Ala Gln Ser Gly Val Met Pro Ile Thr Glu Asn Ser Leu Asp
gcc gaa caa ggt tta gtt ata ccg ccg ttt gtg gag cca cat att cac
Ala Glu Gln Gly Leu Val Ile Pro Pro Phe Val Glu Pro His Ile His
     50
                         55
ctg gac acc acg caa acc gcc gga caa ccg aac tgg aat cag tcc ggc
                                                                   240
Leu Asp Thr Thr Gln Thr Ala Gly Gln Pro Asn Trp Asn Gln Ser Gly
                     70
                                         75
acg ctg ttt gaa ggc att gaa cgc tgg gcc gag cgc aaa gcg tta tta
                                                                   288
Thr Leu Phe Glu Gly Ile Glu Arg Trp Ala Glu Arg Lys Ala Leu Leu
acc cat gac gat gtg aaa caa cgc gca tgg caa acg ctg aaa tgg cag
                                                                   336
Thr His Asp Asp Val Lys Gln Arg Ala Trp Gln Thr Leu Lys Trp Gln
            100
                                105
att gcc aac ggc att cag cat gtg cgt acc cat gtc gat gtt tcg gat
                                                                   384
Ile Ala Asn Gly Ile Gln His Val Arg Thr His Val Asp Val Ser Asp
                            120
gca acg cta act gcg ctg aaa gca atg ctg gaa gtg aag cag gaa gtc
                                                                   432
Ala Thr Leu Thr Ala Leu Lys Ala Met Leu Glu Val Lys Gln Glu Val
    130
                        135
                                            140
gcg ccg tgg att gat ctg caa atc gtc gcc ttc cct cag gaa ggg att
                                                                   480
Ala Pro Trp Ile Asp Leu Gln Ile Val Ala Phe Pro Gln Glu Gly Ile
145
ttg tcg tat ccc aac ggt gaa gcg ttg ctg gaa gag gcg tta cgc tta
                                                                   528
Leu Ser Tyr Pro Asn Gly Glu Ala Leu Leu Glu Glu Ala Leu Arg Leu
                165
                                    170
ggg gca gat gta gtg ggg gcg att ccg cat ttt gaa ttt acc cgt gaa
                                                                   576
Gly Ala Asp Val Val Gly Ala Ile Pro His Phe Glu Phe Thr Arg Glu
                                185
tac ggc gtg gag tcg ctg cat aaa acc ttc gcc ctg gcg caa aaa tac
                                                                   624
Tyr Gly Val Glu Ser Leu His Lys Thr Phe Ala Leu Ala Gln Lys Tyr
                            200
        195
                                                 205
gac cgt ctc atc gac gtt cac tgt gat gag atc gat gac gag cag tcg
                                                                   672
Asp Arg Leu Ile Asp Val His Cys Asp Glu Ile Asp Asp Glu Gln Ser
    210
                        215
                                            220
cgc ttt gtc gaa acc gtt gct gcc ctg gcg cac cat gaa ggc atg ggc
                                                                   720
Arg Phe Val Glu Thr Val Ala Ala Leu Ala His His Glu Gly Met Gly
225
                    230
gcg cga gtc acc gcc agc cac acc acg gca atg cac tcc tat aac ggg
                                                                   768
```

<210> 1

Ala	Arg	Val	Thr	Ala 245	Ser	His	Thr	Thr	Ala 250	Met	His	Ser	Tyr	Asn 255	Gly	
gcg	tat	acc	tca	cgc	ctg	ttc	cgc	ttg	ctg	aaa	atg	tcc	ggt	att	aac	816
Ala	Tyr	Thr	Ser	Arg	Leu	Phe	Arg	Leu	Leu	Lys	Met	Ser	Gly	Ile	Asn	
			260					265					270			
ttt	gtc	gcc	aac	ccg	ctg	gtc	aat	att	cat	ctg	caa	gga	cgt	ttc	gat	864
Phe	Val	Ala	Asn	Pro	Leu	Val	Asn	Ile	His	Leu	Gln	Gly	Arg	Phe	Asp	
		275					280					285				
acg	tat	cca	aaa	cgt	cgc	ggc	atc	acg	cgc	gtt	aaa	gag	atg	ctg	gag	912
Thr	Tyr	Pro	Lys	Arg	Arg	Gly	Ile	Thr	Arg	Val	Lys	Glu	Met	Leu	Glu	
	290					295					300					
tcc	ggc	att	aac	gtc	tgc	ttt	ggt	cac	gat	gat	gtc	ttc	gat	ccg	tgg	960
Ser	Gly	Ile	Asn	Val	_	Phe	Gly	His	Asp	_	Val	Phe	Asp	Pro	-	
305					310					315					320	
	_	ctg		-			_	_			-		_		_	1008
Tyr	Pro	Leu	Gly		Ala	Asn	Met	Leu		Val	Leu	His	Met	_	Leu	
				325					330					335		
	_	tgc	_	_	_				_			_		_		1056
Hls	Val	Cys		Leu	Met	GLY	Tyr	_	GIn	ITe	Asn	Asp	_	Leu	Asn	
			340					345		+			350			1104
		acc			-	-		-	_		_	_	-			1104
ьеи	TTE	Thr 355	птз	птз	ser	Ala	360	1111	ьеu	ASII	ьеu	365	Asp	ıyı	GIY	
att	acc	dcc	aas	aac	agg	acc		ata	att	atc	ata		act	caa	aat	1152
	-	Ala			_	_		_			_	_	-	_		1152
110	370	1114	O L y	11011	501	375	11011	Lea			380	110	1114	Olu	11011	
aaa		gat	aca	cta	cac	cat	caq	att	cca	qta		tat	tca	qta	cat	1200
		Asp		-	-	-	-	-	-	-	-		_	-	_	
385		-			390	_				395	_	-			400	
ggc	ggc	aag	gtg	att	gcc	agc	aca	caa	ccg	gca	caa	acc	acc	gta	tat	1248
Gly	Gly	Lys	Val	Ile	Ala	Ser	Thr	Gln	Pro	Ala	Gln	Thr	Thr	Val	Tyr	
				405					410					415		
ctg	gag	cag	cca	gaa	gcc	atc	gat	tac	aaa	cgt	tga					1284
Leu	Glu	Gln	Pro	Glu	Ala	Ile	Asp	Tyr	Lys	Arg						
			420					425								

<210> 2 <211> 427 <212> PRT <213> Escherichia coli

<400> 2

 Val
 Ser
 Asn
 Asn
 Ala
 Leu
 Gln
 Thr
 Ile
 Ile
 Asn
 Ala
 Arg
 Leu
 Pro
 Gly

 Glu
 Glu
 Gly
 Leu
 Trp
 Gln
 Ile
 His
 Leu
 Gln
 Asp
 Gly
 Lys
 Ile
 Ser
 Ala

 Glu
 Gly
 Leu
 Trp
 Wal
 Met
 Pro
 Ile
 Thr
 Glu
 Asn
 Ser
 Leu
 Asp

 Ala
 Glu
 Glu
 Ser
 Gly
 Val
 Met
 Pro
 Ile
 Thr
 Glu
 Asn
 Ser
 Leu
 Asp

 Ala
 Glu
 Glu
 Leu
 Val
 Ile
 Pro
 Pro
 Pro
 Pro
 Pro
 Pro
 Pro
 Pro
 Asn
 Trp
 Asn
 Ile
 His
 Ile
 His
 Ile
 Ile
 Ile
 Pro
 Pro
 Pro
 Asn
 Trp
 Asn
 Ile
 Ile

```
Ile Ala Asn Gly Ile Gln His Val Arg Thr His Val Asp Val Ser Asp
               120
Ala Thr Leu Thr Ala Leu Lys Ala Met Leu Glu Val Lys Gln Glu Val
                   135
Ala Pro Trp Ile Asp Leu Gln Ile Val Ala Phe Pro Gln Glu Gly Ile
                      155
      150
Leu Ser Tyr Pro Asn Gly Glu Ala Leu Leu Glu Glu Ala Leu Arg Leu
             165
                   170
Gly Ala Asp Val Val Gly Ala Ile Pro His Phe Glu Phe Thr Arg Glu
                           185
Tyr Gly Val Glu Ser Leu His Lys Thr Phe Ala Leu Ala Gln Lys Tyr
                       200
Asp Arg Leu Ile Asp Val His Cys Asp Glu Ile Asp Asp Glu Gln Ser
                    215
Arg Phe Val Glu Thr Val Ala Ala Leu Ala His His Glu Gly Met Gly
                230
                                  235
Ala Arg Val Thr Ala Ser His Thr Thr Ala Met His Ser Tyr Asn Gly
                               250
             245
Ala Tyr Thr Ser Arg Leu Phe Arg Leu Leu Lys Met Ser Gly Ile Asn
          260
                           265
Phe Val Ala Asn Pro Leu Val Asn Ile His Leu Gln Gly Arg Phe Asp
                       280
Thr Tyr Pro Lys Arg Arg Gly Ile Thr Arg Val Lys Glu Met Leu Glu
                           300
                    295
Ser Gly Ile Asn Val Cys Phe Gly His Asp Asp Val Phe Asp Pro Trp
                310
                                  315
Tyr Pro Leu Gly Thr Ala Asn Met Leu Gln Val Leu His Met Gly Leu
             325 330 335
His Val Cys Gln Leu Met Gly Tyr Gly Gln Ile Asn Asp Gly Leu Asn
                           345
Leu Ile Thr His His Ser Ala Arg Thr Leu Asn Leu Gln Asp Tyr Gly
                        360
Ile Ala Ala Gly Asn Ser Ala Asn Leu Ile Ile Leu Pro Ala Glu Asn
                          380
         375
Gly Phe Asp Ala Leu Arg Arg Gln Val Pro Val Arg Tyr Ser Val Arg
                390
                                  395
Gly Gly Lys Val Ile Ala Ser Thr Gln Pro Ala Gln Thr Thr Val Tyr
                        410
             405
Leu Glu Gln Pro Glu Ala Ile Asp Tyr Lys Arg
          420
                           425
<210> 3
<211> 1284
<212> DNA
<213> Artificial sequence
<220>
<223> Description of the artificial sequence: coding for
    cytosine deaminase (codA)
<220>
<221> misc_feature
```

<223> mutation of GTG to ATG start codon for expression

<222> (1)..(3)

in eukaryotic hosts

```
<220>
<221> CDS
<222> (1)..(1281)
<223> coding for cytosine deaminase (codA)
<400> 3
atg tcg aat aac gct tta caa aca att att aac
Met Ser Asn Asn Ala Leu Gln Thr Ile Ile Ass
1 5 10
```

<400> 3																
atg	tcg	aat	aac	gct	tta	caa	aca	att	att	aac	gcc	cgg	tta	cca	ggc	48
Met	Ser	Asn	Asn	Ala	Leu	Gln	Thr	Ile	Ile	Asn	Ala	Arg	Leu	Pro	Gly	
1				5					10					15		
_			_		cag			_	_	_				_	_	96
Glu	Glu	Gly		Trp	Gln	Ile	His		Gln	Asp	Gly	Lys		Ser	Ala	
			20					25					30			
	-				ggc		_				_		_	_	-	144
Ile	Asp		GIn	Ser	Gly	Val		Pro	Ile	Thr	GLu		Ser	Leu	Asp	
		35					40					45				1.00
-	-				gtt		_	-								192
Ala		GIn	GIY	Leu	Val		Pro	Pro	Pne	Val		Pro	HIS	TTE	HIS	
at a	50	200	2.00	a 2 2	200	55	~~~	a 2 2	aaa	226	60	2.2±	a 2 a	+ aa	~~~	240
	_		_		acc Thr	_			_				_			240
65	Asp	1111	TIIL	GIII	70	АІА	СТУ	GIII	PIO	75	пр	ASII	GIII	ser	80	
	cta	+++	caa	aac	att	caa	cac	t aa	acc		cac	222	aca	tta		288
	_		-		Ile	-	_		-		_					200
1111	шец	LIIC	Olu	85	110	Olu	1119	ттр	90	Olu	1119	цуБ	mu	95	шец	
acc	cat	gac	gat		aaa	caa	cac	gca		caa	acq	cta	aaa		cad	336
		_	-		Lys		-	-			_	_			_	
			100		-1-		9	105	1-				110	1		
att	acc	aac		att	caq	cat	ata		acc	cat	atc	σat		tca	gat	384
	-				Gln			_			_	_	_	_	-	
		115	_				120	,				125			-	
gca	acg	cta	act	gcg	ctg	aaa	gca	atg	ctg	gaa	gtg	aag	cag	gaa	gtc	432
Ala	Thr	Leu	Thr	Ala	Leu	Lys	Ala	Met	Leu	Glu	Val	Lys	Gln	Glu	Val	
	130					135					140	_				
gcg	ccg	tgg	att	gat	ctg	caa	atc	gtc	gcc	ttc	cct	cag	gaa	ggg	att	480
Ala	Pro	Trp	Ile	Asp	Leu	Gln	Ile	Val	Ala	Phe	Pro	Gln	Glu	Gly	Ile	
145					150					155					160	
ttg	tcg	tat	ccc	aac	ggt	gaa	gcg	ttg	ctg	gaa	gag	gcg	tta	cgc	tta	528
Leu	Ser	Tyr	Pro	Asn	Gly	Glu	Ala	Leu	Leu	Glu	Glu	Ala	Leu	Arg	Leu	
				165					170					175		
ggg	gca	gat	gta	gtg	ggg	gcg	att	ccg	cat	ttt	gaa	ttt	acc	cgt	gaa	576
Gly	Ala	Asp	Val	Val	Gly	Ala	Ile	Pro	His	Phe	Glu	Phe	Thr	Arg	Glu	
			180					185					190			
tac	ggc	gtg	gag	tcg	ctg	cat	aaa	acc	ttc	gcc	ctg	gcg	caa	aaa	tac	624
Tyr	Gly		Glu	Ser	Leu	His	_	Thr	Phe	Ala	Leu		Gln	Lys	Tyr	
		195					200					205				
_	_			-	gtt		-	-			-	-		_	_	672
Asp	_	Leu	Ile	Asp	Val		Суз	Asp	Glu	Ile	_	Asp	Glu	Gln	Ser	
	210					215					220					
_		_	_		gtt	-	_	_				_		_		720
_	rne	val	GIU	ınr	Val	АІА	АІА	ьeu	АІА		HlS	GLU	СТА	мet		
225					230					235					240	7.60
	_	-		-	agc			_	-	_						768
нта	Arg	val	ınr		Ser	п15	ınr	ınr		мет	nlS	ser	туr		стλ	
~~~	+ 2+	200	t c :	245	at~	++~	ccc	++~	250	200	a+~	+ ~ ~	~~+	255	226	216
				_	ctg		_	_	_		_					816
AId	тУТ	TIIL	260	v1.	Leu	rne	v1.	eu	ьeu	туз	riet	Set	270	тте	ASH	
			200					200					2/0			

ttt	gtc	gcc	aac	ccg	ctg	gtc	aat	att	cat	ctg	caa	gga	cgt	ttc	gat	864
Phe	Val	Ala	Asn	Pro	Leu	Val	Asn	Ile	His	Leu	Gln	Gly	Arg	Phe	Asp	
		275					280					285				
acg	tat	cca	aaa	cgt	cgc	ggc	atc	acg	cgc	gtt	aaa	gag	atg	ctg	gag	912
Thr	Tyr	Pro	Lys	Arg	Arg	Gly	Ile	Thr	Arg	Val	Lys	Glu	Met	Leu	Glu	
	290					295					300					
tcc	ggc	att	aac	gtc	tgc	ttt	ggt	cac	gat	gat	gtc	ttc	gat	ccg	tgg	960
Ser	Gly	Ile	Asn	Val	Cys	Phe	Gly	His	Asp	Asp	Val	Phe	Asp	Pro	Trp	
305					310					315					320	
tat	ccg	ctg	gga	acg	gcg	aat	atg	ctg	caa	gtg	ctg	cat	atg	ggg	ctg	1008
Tyr	Pro	Leu	Gly	Thr	Ala	Asn	Met	Leu	Gln	Val	Leu	His	Met	Gly	Leu	
				325					330					335		
cat	gtt	tgc	cag	ttg	atg	ggc	tac	aaa	cag	att	aac	gat	ggc	ctg	aat	1056
His	Val	Cys	Gln	Leu	Met	Gly	Tyr	Gly	Gln	Ile	Asn	Asp	Gly	Leu	Asn	
			340					345					350			
tta	atc	acc	cac	cac	agc	gca	agg	acg	ttg	aat	ttg	cag	gat	tac	ggc	1104
Leu	Ile	Thr	His	His	Ser	Ala	Arg	Thr	Leu	Asn	Leu	Gln	Asp	Tyr	Gly	
		355					360					365				
att	gcc	gcc	gga	aac	agc	gcc	aac	ctg	att	atc	ctg	ccg	gct	gaa	aat	1152
Ile	Ala	Ala	Gly	Asn	Ser	Ala	Asn	Leu	Ile	Ile	Leu	Pro	Ala	Glu	Asn	
	370					375					380					
ddd	ttt	gat	gcg	ctg	cgc	cgt	cag	gtt	ccg	gta	cgt	tat	tcg	gta	cgt	1200
Gly	Phe	Asp	Ala	Leu	Arg	Arg	Gln	Val	Pro	Val	Arg	Tyr	Ser	Val	Arg	
385					390					395					400	
ggc	ggc	aag	gtg	att	gcc	agc	aca	caa	ccg	gca	caa	acc	acc	gta	tat	1248
Gly	Gly	Lys	Val		Ala	Ser	Thr	Gln		Ala	Gln	Thr	Thr		Tyr	
				405					410					415		
ctg	gag	cag	cca	gaa	gcc	atc	gat	tac	aaa	cgt	tga					1284
Leu	Glu	Gln		Glu	Ala	Ile	Asp	_	Lys	Arg						
			420					425								

<210> 4

<211> 427

<212> PRT

<213> Artificial sequence

<220>

<223> Description of the artificial sequence: coding for cytosine deaminase (codA)

<400> 4

Met Ser Asn Asn Ala Leu Gln Thr Ile Ile Asn Ala Arg Leu Pro Gly 5 10 Glu Glu Gly Leu Trp Gln Ile His Leu Gln Asp Gly Lys Ile Ser Ala 25 Ile Asp Ala Gln Ser Gly Val Met Pro Ile Thr Glu Asn Ser Leu Asp 40 Ala Glu Gln Gly Leu Val Ile Pro Pro Phe Val Glu Pro His Ile His 55 60 Leu Asp Thr Thr Gln Thr Ala Gly Gln Pro Asn Trp Asn Gln Ser Gly Thr Leu Phe Glu Gly Ile Glu Arg Trp Ala Glu Arg Lys Ala Leu Leu 85 90 Thr His Asp Asp Val Lys Gln Arg Ala Trp Gln Thr Leu Lys Trp Gln 105 110

Ile Ala Asn Gly Ile Gln His Val Arg Thr His Val Asp Val Ser Asp

		115					120					125			
Ala	Thr	Leu	Thr	Ala	Leu	Lys	Ala	Met	Leu	Glu	Val	Lys	Gln	Glu	Val
	130					135					140				
Ala	Pro	Trp	Ile	Asp	Leu	Gln	Ile	Val	Ala	Phe	Pro	Gln	Glu	Gly	Ile
145					150					155					160
Leu	Ser	Tyr	Pro	Asn	Gly	Glu	Ala	Leu	Leu	Glu	Glu	Ala	Leu	Arg	Leu
				165					170					175	
Gly	Ala	Asp	Val	Val	Gly	Ala	Ile	Pro	His	Phe	Glu	Phe	Thr	Arg	Glu
			180					185					190		
Tyr	Gly	Val	Glu	Ser	Leu	His	Lys	Thr	Phe	Ala	Leu	Ala	Gln	Lys	Tyr
		195					200					205			
Asp	Arg	Leu	Ile	Asp	Val	His	CA2	Asp	Glu	Ile	Asp	Asp	Glu	Gln	Ser
	210					215					220				
Arg	Phe	Val	Glu	Thr	Val	Ala	Ala	Leu	Ala	His	His	Glu	Gly	Met	Gly
225					230					235					240
Ala	Arg	Val	Thr	Ala	Ser	His	Thr	Thr	Ala	Met	His	Ser	Tyr	Asn	Gly
				2.											